Technical Report 12-004

Data Collection Challenges: Lessons Learned from the Czech Capital Markets

Margaret Moten
Evan Szablowski
Kelley Cassidy
Jack Balstad

U.S. Military Academy, West Point NY

February 2012

United States Military Academy
Network Science Center

**U.S. Military Academy
Network Science Center**

Authorized and approved for distribution:

(signature)   *(signature)*

BG (R) CHRIS ARNEY, Ph.D.   **COL JOHN M GRAHAM, Ph.D.**
Network Science Chair   **Director, Network Science Center**

Technical review by

Kathryn Coronges, Ph.D., Department of Behavioral Sciences and Leadership, U.S. Military Academy
Daniel Evans, Network Science Center, U.S. Military

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YYYY)* 13-02-2012	2. REPORT TYPE Technical Report	3. DATES COVERED *(From - To)* Mar-July 2011

4. TITLE AND SUBTITLE Data Collection Challenges: Lessons Learned from the Czech Capital Markets	5a. CONTRACT NUMBER n/a
	5b. GRANT NUMBER n/a
	5c. PROGRAM ELEMENT NUMBER n/a
6. AUTHOR(S) Margaret Moten, Evan Szablowski, Kelley Cassidy, and Jack Balstad	5d. PROJECT NUMBER ARO NetSci 02
	5e. TASK NUMBER n/a
	5f. WORK UNIT NUMBER n/a

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Network Science Center, U.S. Military Academy 601 Cullum Road, Thayer Hall Room 119 West Point, NY 10996	8. PERFORMING ORGANIZATION REPORT NUMBER n/a

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S.Army Research Organization Triangle Park, NC	10. SPONSOR/MONITOR'S ACRONYM(S) USMA NSC
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) 12-004

12. DISTRIBUTION / AVAILABILITY STATEMENT

Unlimited Distribution

13. SUPPLEMENTARY NOTES
The views expressed in this thesis are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

14. ABSTRACT
The Network Science Center's Frontier Markets research project is developing methodologies to classify Capital Market networks. A research team consisting of two researchers and three cadets from the Network Science center visited the Czech Republic from 12-19 March 2011. The purpose of the trip was to conduct research on the capital market network in Prague, validate-initial network results, and identify individuals, organizations and roles that should be incorporated into subsequent network typologies. The team has developed network models of three Frontier Markets: Tanzania, Ghana, and Trinidad & Tobago. One of the team's research goals is to construct a quantitative "vertical comparison" between our Frontier Markets and an Emerging Market for a network perspective.

15. SUBJECT TERMS
Social Network Analysis, Network Science, Economics, Capital Markets, Frontier Markets

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Tish Torgerson
a. REPORT UNCLASSIFIED	b. ABSTRACT UNCLASSIFIED	c. THIS PAGE UNCLASSIFIED	UL	7	19b. TELEPHONE NUMBER *(include area code)* 845-938-0804

Data Collection Challenges: Lessons Learned from the Czech Capital Markets

Margaret Moten
Evan Szablowski
Kelley Cassidy
Jack Balstad

U.S. Military Academy, West Point NY

ACKNOWLEDGEMENT

July 8, 2011

Data Collection Challenges: Lessons Learned from the Czech Capital Markets

Margaret Moten, Evan Szablowski, Kelley Cassidy, and Jack Balstad

BACKGROUND

The Network Science Center's Frontier Markets research project is developing methodologies to classify capital market networks[1]. As part of this effort, a research team consisting of two researchers and three cadets from the Network Science Center visited the Czech Republic from 12 - 19 March, 2011. The purpose of the trip was to conduct research on the capital market network in Prague, validate initial network results, and identify individuals, organizations and roles that should be incorporated into subsequent network typologies. The team has developed network models of three frontier markets: Tanzania, Ghana, and Trinidad & Tobago. One of the team's research goals is to construct a quantitative vertical comparison between our frontier markets and an emerging market from a network perspective. We selected the Czech Republic as the emerging market because they liberalized their markets and privatized many government-owned industries at approximately the same time as Tanzania and Ghana. The availability of data and the use of English in business were also major determining selection factors.

METHODOLOGY

Prior to the trip, the team built a social network model of the Czech capital markets using information about firms, companies, and organizations that is readily available on the Internet. The research team identified links between individuals and organizations based on shared affiliations. The team initially analyzed the stock exchange and its leadership as well as the major brokers that operate on the exchange.

---

[1] For more information on the team's methodology, please see the team's project summary paper at www.netscience.usma.edu.

Then they examined the government organizations that influence and regulate the capital market. This paper is a reflection on the challenges of collecting this type of data authored by cadet researchers from the United States Military Academy.

In order to develop an initial social network model of the Czech capital markets, the research team began with the organization typically considered the hub of an economy: the stock exchange. Cadet researchers collected data on the directors and managers of the Prague Stock Exchange, as well as, key personnel at the Ministry of Finance and major banks and broker/dealers operating on the exchange. Researchers documented each individual's current organization and position, past organizational affiliations, and boards of directors on which the person served. Additional information compiled included nationality, universities attended and degrees granted, and social and professional organizations. Researchers uploaded the data collection spreadsheet into the ORA[2] network analysis software to build the initial network depicting the interrelationships among individuals and organizations based on shared affiliations. Using matrix algebra functionality in ORA, the team also generated a network that connected organizations to other organizations through individuals and a network depicting how people are connected to other people through the organizations and institutions with which they are associated.

ORA generated myriad network indications (such as density, characteristic path length, and betweenness and closeness centrality) on these initial networks highlighting which individuals and organizations were the most influential and interconnected. These networks allowed researchers to detect the key actors and organizations for further research. Specifically, individuals and organizations high in network centrality were identified as important in the emerging markets and targeted for on-the-ground interviews.

---

[2] Carley, Kathleen M. (2001-2010). ORA: Organizational Risk Analyzer v.2.02. [Network Analysis Software]. Pittsburgh: Carnegie Mellon University.

INITIAL NETWORK ANALYSIS

The following figures graphically depict the relationships among agents and organizations in the Czech Capital Market Network and reveal key nodes, clusters, bridges, and isolates. Figure I shows the shared affiliation network of agents and organizations in the initial cadet model. The red nodes represent agents; the green nodes are organizations.

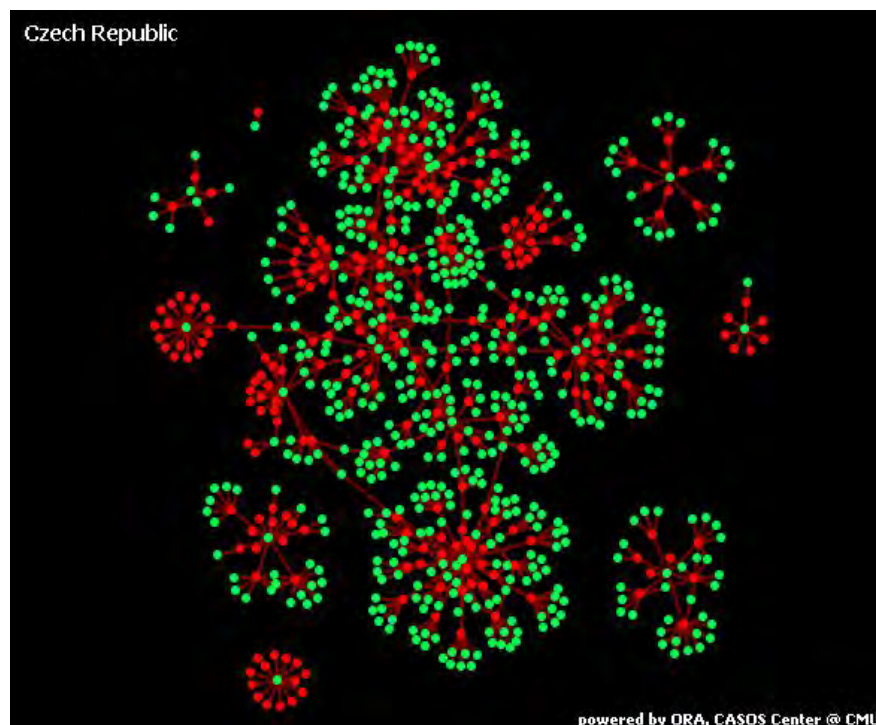**Figure I: Initial Agent - Organization Model Results**

Figure II depicts how individuals (agents) are connected to other individuals through organizations based on affiliations with the same organizations or institutions. Figure III shows how organizations are connected to other organizations based on their affiliations with individuals.

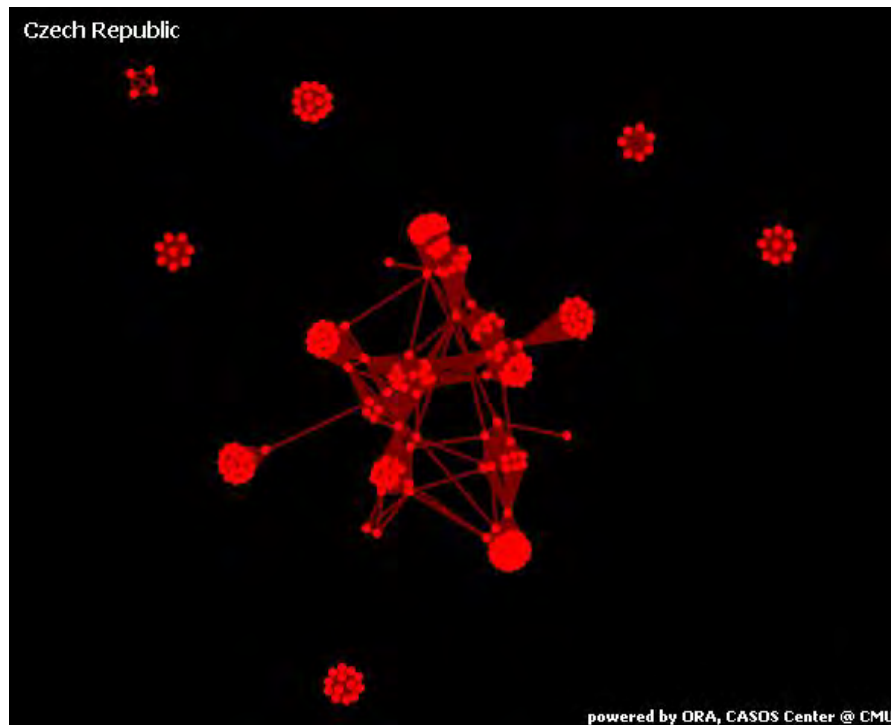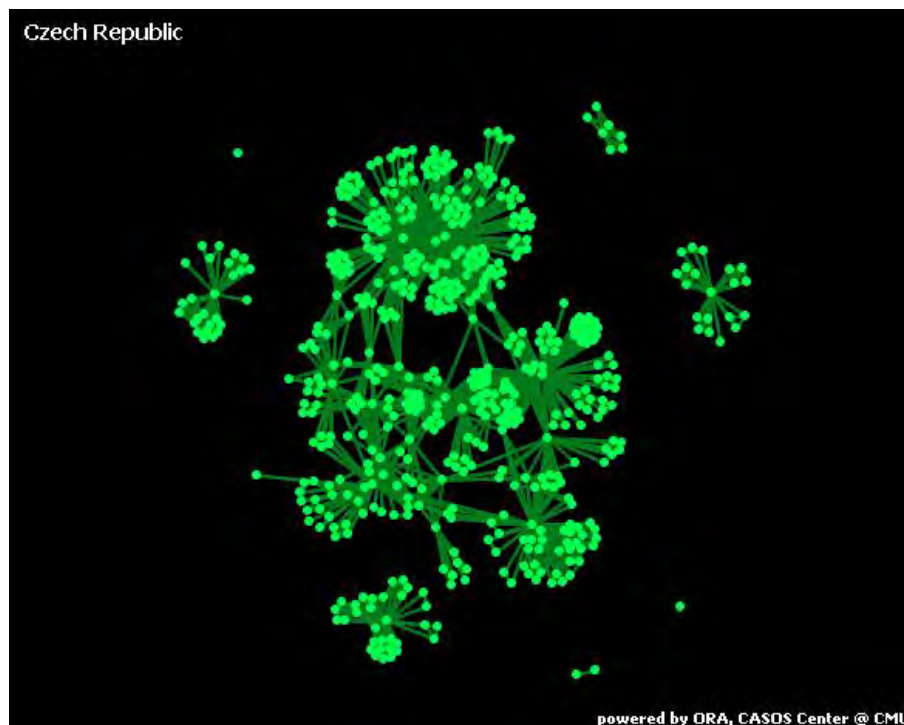## Figure II: Initial Agent - Agent Model Results

Figure III: Initial Organization - Organization Model Results

INTERVIEWS

On the ground research in the Czech Republic provided additional insights that will be incorporated into future network typologies. Researchers interviewed individuals whom the initial capital market network revealed to be key nodes in the network. During the visit, the team met with the Economics Counselor at the U.S. Embassy (to obtain background information), and representatives from the Financial Markets Department of the Czech Ministry of Finance, the Banking Supervisory Department of the Czech National Bank, four of the five largest brokerage firms that operate on the Prague Stock Exchange, and two different private equity firms that operate in the Czech Republic (one focusing on small and medium enterprises in the Czech Republic and one larger, international firm that focuses primarily on Eastern Europe). The interviews covered topics ranging from the development and operation of the Prague Stock Exchange to the banking and regulatory environment.

LESSONS LEARNED

The team learned that the Prague Stock Exchange is not the most central organization in the Czech financial system. For many reasons, including a risk-averse, non-entrepreneurial culture and problems with the initial privatization of government-owned firms after the collapse of Communism, few Czech citizens are heavily invested in the stock market. The Prague Stock Exchange has few actively traded securities and most sophisticated Czech shareholders invest in securities and funds abroad. Many Czech firms remain privately held and few incentives exist to entrepreneurs to pursue initial public offerings.

The team also faced many challenges with data collection and research. Information availability varied widely among the entities researched. Some websites provided extensive biographies for their executives, while others listed very little or nothing at all. Some individuals may appear to be extremely influential relative to their peers because they have chosen to publish extensive background information. Conversely, influential individuals may prefer to remain anonymous posting little

personal information on the Internet. Consequently, the initial network may contain structural holes if an undocumented relationship existed among individuals or organizations, due to either errors of omission or commission.

Also, many firms may not update their websites frequently, so researchers conducted additional assessments to validate individual résumé data. However, the network may still be incomplete if an individual has become a member of another board, completed a university degree, or joined a professional organization since his or her résumé was posted. Data collection is further complicated by the difficulty of keeping the dataset up to date. Short of developing a web crawler or checking corporate Internet sites regularly, it is difficult to know when a key person changes institutions or a new individual joins a key organization. It is quite possible that by the time this research is completed, the network typologies will have lost a significant level of their accuracy in reflecting the Czech capital market. As a result, these static models are not very sensitive to changes among the economic actors and entities within the capital market.

Perhaps the greatest limitation in the network arises from a key assumption: if two people are associated with an organization or institution, they have a significant link or connection. In reality, this simplifying assumption causes the model to overstate some relationships. For example, the Prague capital market network shows links between two people who went to the same university; however, universities are large organizations and student ages and fields of study vary such that two individuals attending the same university may never have met. Likewise, individuals may have worked for the same organization or served on boards of directors at different times making it difficult to know if an actual link exists between them.

Fortunately, the ORA software allows links to be weighted in order to accommodate these types of limitations. For example, since accurately reflecting university affiliations is problematic, researchers can weight the links less heavily than the weight assigned to current organizations or social clubs. If two individuals matriculated from the same university during the same time period with similar degrees,

the links could also be weighted more heavily. These weightings would, in effect, ensure weak connections have less influence on the network's statistical measures, while strong connections would have greater influence on network structure and statistics. Adjusting link weights would produce a more accurate model of the Czech capital market network.

FUTURE RESEARCH

In order to more accurately reflect the capital market network, researchers will expand the data set to include influential Czech's involved in the capital markets who were included in the initial model, the directors and executives of the companies with which they are affiliated, and the key individuals at non-financial publicly-traded companies, consulting groups, professional groups (for example, the bar association, CPA firms, and CFA members), and industry trade groups. Based on these revisions, the team will publish a paper that quantitatively compares the initial model against the revised model. The team will use this comprehensive model as part of its larger effort to compare and contrast frontier capital markets and emerging capital markets.